12-2020

Classification and Prediction Models for Natural Streamflow Regimes in the Arid Southwestern USA

Angela M. Merritt
Utah State University

CLASSIFICATION AND PREDICTION MODELS FOR NATURAL STREAMFLOW REGIMES IN THE ARID SOUTHWESTERN USA

by

Angela M. Merritt

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Watershed Science

Approved:

Charles Hawkins, Ph.D. Major Professor	Belize Lane, Ph.D. Committee Member
Peter Wilcock, Ph.D. Committee Member	D. Richard Cutler; Ph.D. Interim Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2020

ABSTRACT

Classification and Prediction Models for Natural Streamflow

Regimes in the Arid Southwestern USA

by

Angela M. Merritt, Master of Science

Utah State University, 2020

Major Professor: Dr. Charles Hawkins
Department: Watershed Science

Understanding how natural variation in flow regimes influences stream ecosystem structure and function is critical to the development of effective stream management policies and actions. Spatial variation in flow regimes is well understood for stream reaches in mesic regions, but a more robust characterization of flow regimes in arid regions is needed, especially to support biological monitoring and assessment programs. Methods are specifically needed that can accurately predict the flow regime expected at ungauged reaches. We used long-term (41 y) records of mean daily flow from 287 stream reaches in the arid western USA to develop and compare several alternative classifications. We also evaluated how accurately we could predict flow regime classes from topographic, soil, and climatic data. Over the 41-y record examined (1972 – 2013), the 287 stream reaches varied continuously from being always wet (perennial) to being dry most days. As a robust characterization of flow regimes in the southwest USA:

1) We identified 3 hierarchical levels of classification and interpreted the most resolved 5-group classification to include ephemeral, nonperennial snowmelt-driven,

perennial snowmelt-driven, nonperennial rain-driven, and mixed perennial/nonperennial, rain-driven flow regime classes.

2) We created a second set of 4 classifications based on the percentages of zero flow days (ZFD) and years with zero flows (ZFY).

We built random forest models to predict streamflow in variable latitudes, longitudes, and elevations. We used classification models to predict class membership and added 2 random forest regression models to directly predict the ZFD and ZFY for streamflow records. Ephemeral and perennial stream reaches were predicted with less error than stream reaches with intermediate nonperennial days or years. The regression models explained ~ 50% of the variation in both percent of ZFD and ZFY. Water resource managers of arid regions and sub-regions should select the desired classification in order to determine expected streamflow across each watershed.

(60 pages)

PUBLIC ABSTRACT

Classification and Prediction Models for Natural Streamflow

Regimes in the Arid Southwestern USA

Angela M. Merritt

Understanding how natural variation in flow regimes influences stream ecosystem structure and function is critical to the development of effective stream management policies and actions. Spatial variation in flow regimes is well understood for stream reaches in mesic regions, but a more robust characterization of flow regimes in arid regions is needed, especially to support biological monitoring and assessment programs. Methods are specifically needed that can accurately predict the flow regime expected at ungauged reaches. We used long-term (41 y) records of mean daily flow from 287 stream reaches in the arid western USA to develop and compare several alternative classifications. We also evaluated how accurately we could predict flow regime classes from topographic, soil, and climatic data. Over the 41-y record examined (1972 – 2013), the 287 stream reaches varied continuously from being always wet (perennial) to being dry most days. We explored 5 hierarchical levels of classification and interpreted the 5-group classification to include ephemeral, nonperennial snowmelt-driven, perennial snowmelt-driven, nonperennial rain-driven, and mixed perennial/nonperennial, rain-driven flow regime classes. We created a second set of 4 classifications based on the percentages of days and years with zero flows. We then built random forest classification models to predict class membership, in addition to 2 random forest regression models to directly predict the mean percent of days in a year with zero flow and the number of years with zero flow. Ephemeral and perennial stream reaches were predicted with less error than stream reaches with intermediate nonperennial days or years. The regression models explained ~ 50% of the variation in both percent of zero flow days in a year and percent of zero flow years. These models would predict flow regimes at ungauged reaches in Arizona, identifying ephemeral flow regimes. Maps based on these predictions were generally consistent with qualitative expectations of how flow regimes varied spatially across the state, but larger Arizona stream reaches were predicted with more error than smaller stream reaches. These results represent a promising step toward more effective stream assessment and management in arid regions.

ACKNOWLEDGMENTS

I thank my major advisor, Dr. Charles Hawkins, and my committee members, Dr. Belize Lane and Dr. Peter Wilcock, all of whom guided me as I took on a very new subject matter in my academic career. I would also like to thank my primary lab mate, Donald Benkendorf, who discussed many primary methods and analyses with me. Christian Perry, Adam Fisher, Brennan Bean, Emily Burchfield, and Ryan Hill discussed much code, many statistical analyses, and several visualizations, for which I owe great thanks. I would also like to thank the USU graduate community, students, staff, and faculty, especially those in the Watershed Sciences Department, who guided and supported me during my tenure at the university. Kendall Becker was exceptionally helpful as a writing coach and counselor.

I would like to especially thank Patti Spindler and Susan Jackson for guidance and encouragement regarding the need for and applicability of these models. Patti was particularly supportive as my field guide.

I would like to thank my mother, who continues to support all my pursuits, and lastly my pops, who did not get to see the final product. Dad, you were immensely helpful in this and all things life. I love you and mom bunches.

Angie Merritt

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Water resource managers need to characterize streamflow regimes to support the biological monitoring and assessment of stream ecosystems [1-4]. Naturally occurring differences in annual and interannual streamflow patterns are important drivers of variation in aquatic biodiversty and ecological processes [5-9]. Characterizing the full range of natural (aka reference) streamflow variation throughout a region is a critical first step when developing assessment tools to evaluate the biological condition of streams. The aquatic life expected to occur in an assessed stream is typically estimated from the biota observed at a set of minimally or least-altered reference sites that best match the environmental conditions that would naturally occur at the assessed site [3, 10, 11]. In environmentally heterogeneous regions, a particularly important challenge is to identify the range of specific environmental and biological conditions that represent different reference states [10]. This challenge is especially acute for nonperennial streams, which have been understudied relative to perennial streams and for which we have a poor understanding of their physical, chemical, and biological diversity [1, 4, 12]. In this paper, we define nonperennial streams as those that cease flowing for at least one day over some defined number of years of record [see Busche et al. 13]. In arid regions, stream networks are typically dominated by nonperennial stream reaches [14, 15], although perennial stream reaches may occur as well.

In general, the aquatic biota that inhabits nonperennial streams is distinct from that inhabiting perennial streams [16-19], and in some ways streams in arid regions are even more critical to overall landscape health than their mesic counterparts [15, 20-22]. However, it is not yet clear how variable or predictable nonperennial streams are from one another in terms of either their hydrologic regimes or the biota they support [4, 17-19, 23, 24]. Water resource managers need to more fully characterize the diversity of flow regimes that occur in these regions, and they need to be able to predict and map where they occur [20-22, 25-27]. Our lack of understanding of the hydrologic heterogeneity that exists across streams in arid regions, and the extent to which that heterogeneity is linked to naturally occurring variation in biota, currently limits the development and application of robust bioassessment tools for streams throughout arid regions [1, 4, 18, 24, 28, 29]. Accurately classifying the hydrological regimes of the 1,000s of ungauged streams in arid regions is a critical precursor to the development and application of bioassessment programs in these regions [15, 25, 28-31].

Perhaps the most striking difference between streams in arid and mesic regions is the extent to which arid-region streams experience drying. Over 90% of the stream network in an arid region may be nonperennial, and the degree to which streams are nonperennial can vary greatly from one or a few days of zero flow per year to most days having no surface flow over numerous years [13-15, 20, 21]. Some nonperennial streams, typically in mountainous arid regions, are nearly perennial, with few or no zero-flow days in at least some years and seasonal snowmelt or storm-driven streamflow pulses [20, 21, 32-34]. The flow regimes of these nearly perennial streams are influenced by differences in elevation and resulting differences in the type and timing of precipitation, as observed in perennial streams [34-40]. These flow regimes are also often influenced by more extreme seasonal differences in snow cover and extended periods of drought than observed in the majority of perennial or near perennial streams [15, 20, 21, 34, 41]. Other streams in arid regions can have ephemeral flow regimes, characterized by short-duration, high-peak flow events interrupting extended periods of zero flow associated with very low or no baseflow throughout the year [14, 15, 20, 21, 32, 33]. These flow regimes are influenced by high evapotranspiration rates, variable drought cycles, and episodic rain events interrupting droughts [34-36].

Researchers and managers have previously classified streamflow regimes into discrete groupings as a way to generalize about how streams differ in their flow characteristics, quantify their diversity, and predict biotic responses to flow [6, 9, 37-40, 42]. Most previous flow classifications have used hierarchical cluster analysis to identify classes. These analyses are ideally based on a large number of sites with continuous, long-term (>20 years) records that robustly characterize the observed daily streamflow patterns through space and time [37-40]. Unfortunately, arid regions generally have limited gauge records, and those streams that are gauged often have short or incomplete records [37, 43-45]. Use of short-term or incomplete records can result in inaccurate and unrepresentative characterization of the actual flow regimes that exist. These types of data limitations can introduce large uncertainties in both flow regime classifications [45, 46] and their prediction [25]. These data limitations have compromised the extent to which we have been able to accurately characterize reference conditions for streams in semi-arid and arid regions [15, 25, 28, 29].

Even if an accurate classification of the variable flow regimes that occur in arid-regions streams existed, predicting what class ungauged streams belong to may be more challenging in arid regions than mesic regions because of our limited understanding of some of the physicoclimatic controls on stream flows in these regions. Many watershed attributes previously used to predict perennial streamflow regimes [37-40] may be less influential in nonperennial streams [26, 27], and additional physicoclimatic attributes, such as evapotranspiration [20, 41], snowcover [47], or streambed topography, may be needed to better predict the occurrence and frequency of zero flow conditions [31, 35, 36, 48-50]. Furthermore, low correlation and non-linear dependencies between rainfall and runoff in arid regions [51] may fundamentally constrain our ability to predict flow regimes of arid-region streams. However, predictions of stream classes at ungauged streams are critical to the success of bioassessment programs because the vast majority of assessments are conducted on streams that lack streamflow gauges.

These classification and prediction challenges are exemplified in the most arid regions of the southwestern USA. For example, only 25 US Geological Survey (USGS) reference streamflow gauge records exist across Arizona [52]. Seventeen of these records start in the mid 1900s, but they have significant gaps or cease completely after 15-20 years, most notably after 1985 [53]. Furthermore, these streamflow records are especially sparse and incomplete for nonperennial reaches [37, 39, 53], a pattern that is true for most extremely arid regions [1, 14]. The limited availability of streamflow records in the southwestern USA presents a specific challenge for both the identification of flow regime classes and their spatial prediction [37, 25]. Here, we present an approach to augment the spatial and temporal availability of streamflow data in an arid region with limited gauges and then use these extended streamflow records to develop and evaluate (1) alternative streamflow regime classifications applicable to the arid southwestern USA and (2) empirical models to predict the streamflow regimes of ungauged streams with publicly available geospatial data.

MATERIALS AND METHODS

2.1 General approach

To address our objectives, we used existing streamflow records from reference-quality streams located throughout the arid southwestern USA (Figure 1). The streamflow dataset obtained from the USGS *gagesII* reference dataset consisted of either full or partial records for 287 locations in the western USA with minimally impaired flow [52]. The workflow required to complete analyses consisted of several steps (Figure 2). We first identified stream gauge records (n = 90) with continuous (or nearly continuous) records that spanned at least 41 years (1972-2013). We then compiled flow records from gauges in our study area with abbreviated or discontinuous data that were collected during this time period (n = 197). We developed Random Forest (RF) regression models in R [54, 55, 56] to predict the full record of mean daily flows for each of the 197 reaches with partial flow records from the data observed at the 90 gauges with continuous records. These model-generated streamflow data were then used to extract a variety of site-specific flow metrics or statistics. These metrics represented components of either streamflow magnitude, frequency, duration, timing, and rate-of-change [6, 39] or components of the frequency, duration, and timing of zero flows [36, 37]. We then used hierarchical cluster analysis to identify groups of streams that differed in one or more flow metrics. We also calculated the number of zero flow days that occurred in each record and summarized these data as 1) the portion of a record with zero flow days and 2) the portion of the 41 water years containing one or more zero flow days. From these summary data, we developed additional classification schemes based on different threshold values of 1) the percent of zero flow days (ZFD) and 2) the percent of years with one or more zero flow days (ZFY). We then developed RF models to predict class membership for each of these classifications. We also used RF to model the continuous variation in percent ZFD and percent ZFY observed in the records. For each of the models, we evaluated how well we could predict flow regimes at ungauged streams from readily available or derived geospatial data related to landscape and climate conditions (see Appendix A).

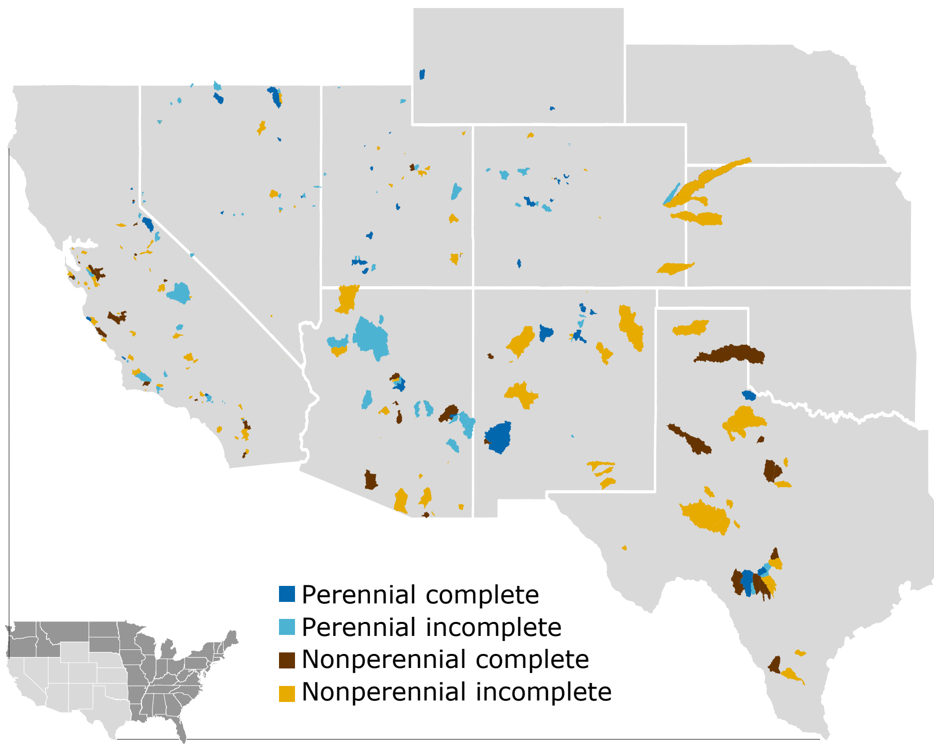

Figure 1. Map of the 287 basins with complete (dark blue, dark brown) and partial (light blue, light brown) USGS gagesII reference streamflow data. All onperennial streams are in brown and all perennial streams are in blue.

2.2 Study region

The southwestern USA was used as the study region for our analyses (Figure 1). This region extends from the southernmost point of the California coastline to the southernmost point of the Texas coastline and up to 1500 km north of the USA-Mexico border. As a whole, the USA has one of the largest and longest datasets of gauged streamflow in the world [43, 44]. The southwest USA region has the best representation

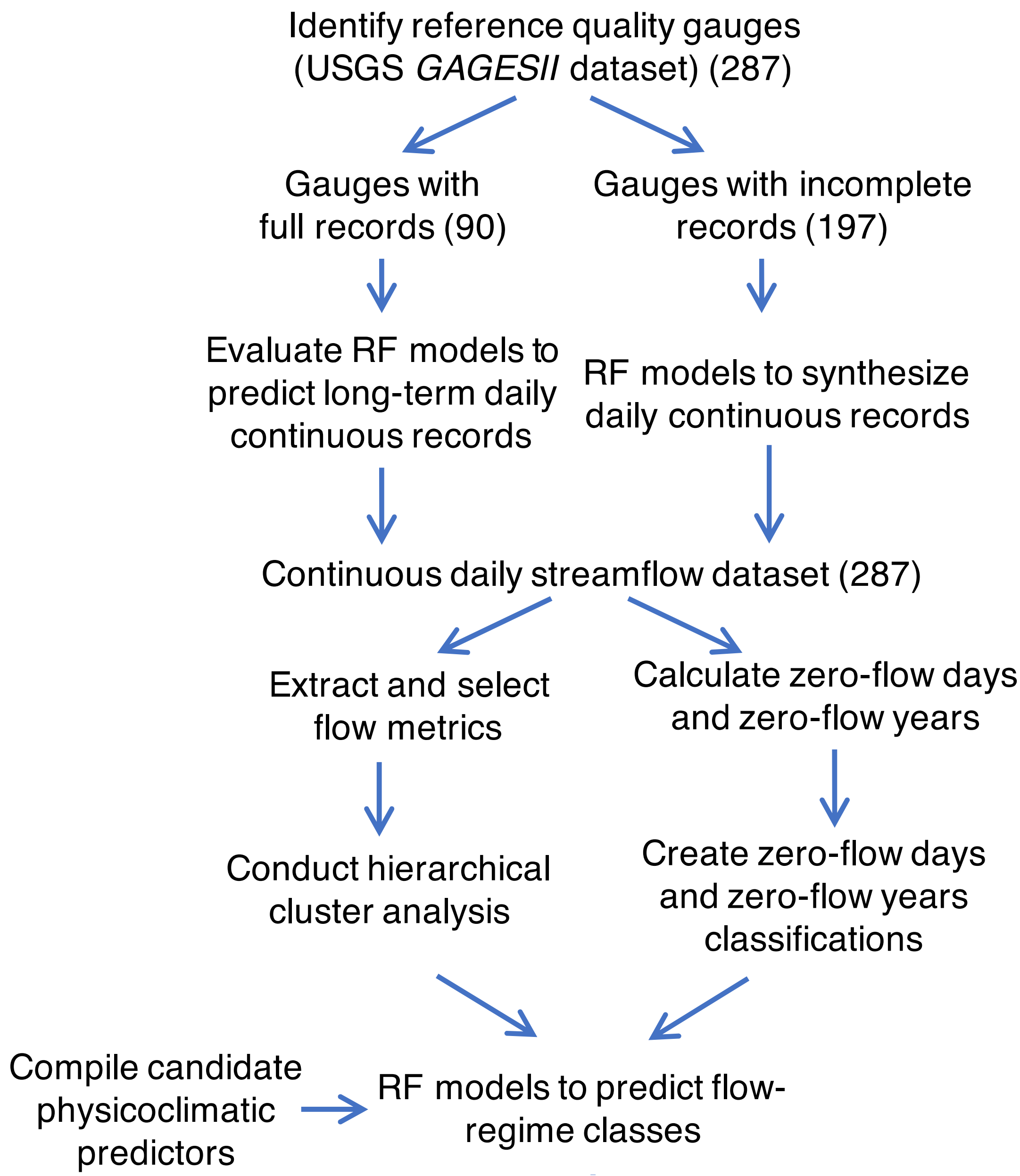

Figure 2. Workflow describing data compilation, pre-analysis data manipulation, classification, modeling, and mapping.

of gauged minimally impaired streams of any arid region in the world [57]. Furthermore, the region spans 22 Level III Ecoregions [58] and presents the largest percentage of nonperennial streams with unimpeded flows across the entire USA [15, 37, 52, 53]. Within the study region, multiple stream gauges have records predating 1920 [53].

2.3 Streamflow data collection

We identified 287 gauged streams with unimpeded flows within our study region. Ninety streams had 41 water-years of mean daily flow measurements (1972-2013), and 197 streams had 10 or more years of missing data (Table 1; Figure 1). Forty-five of the 90 complete streams were perennial (no zero-flow days in the period of record), and 45 streams were nonperennial. Of the remaining 197 gauges with partial records, 167 were nonperennial with between 2% to over 99% zero-flow days. Of the 167 nonperennial streams, 80 had records with less than 20% zero-flow days and 87 had records with more than 20% zero-flow days (Table 1).

2.4 Creating complete streamflow records for all gauged streams

We developed random forest (RF) models [53-55] to predict missing daily flow values at the 197 streams with incomplete data from the daily flow values measured at the 90 streams with complete records. These models predicted daily flow with pseudo-r^2 values ranging between 0.42 and 0.99. Only 3 of these 90 models explained less than 80% of the variation in observed values. Furthermore, models for 240 of the full set of 287 streams accounted for over 90% of the variation in observed flow values. Another 17 models accounted for between 70% and 90% of observed variation in mean daily flows, and the remaining 20 models accounted for less than 70% of the variation in daily flows.

With one exception, the models with the lowest pseudo-r^2 values (-2 to 70%) were those for nonperennial streams (Table 1). Of the 20 records for which models explained less than 70% of observed variance in mean daily flows, 19 were from nonperennial streams with high percentages of zero-flow days. The exception was a perennial stream with nearly constant flow.

Table 1. Totals for initial categories of the streamflow gauge data collected at gagesII locations across the Southwest USA study region and the range of variation explained by the RF models for each category.

Type of record	Number	Variance explained (%)
Perennial	120	45.3–99.6
Nonperennial	167	-02.3–99.5
Complete	90	41.8–99.6
Nonperennial (NP)	45	41.8–99.5
NP *<20*% zero flow*	*21	86.9–99.5
NP *>*20% zero flow*	*24	41.8–98.9
Perennial	45	80.0–99.6
Partial	197	-02.3–99.6
Nonperennial (NP)	122	67.0–99.3
NP *<20*% zero flow*	*59	67.0-99.3
NP *>*20% zero flow*	*63	-02.3–99.0
Perennial	75	45.3–99.6

All regression models can predict negative values at low values. To avoid negative values in subsequent analyses, we adjusted predicted daily flow values to maintain the same proportion of zero-flow values as in the raw data for each stream. We first calculated the percentage of days with mean daily flow equal to zero based on the original data, regardless of whether records were complete or partial. We then identified the predicted daily flow value associated with this percentage for each stream record. Next, we identified all occurrences of mean daily flow at or below this stream-specific threshold. Finally, for each predicted record, we replaced all daily flows at or below their stream-specific threshold flow values with zero.

2.5 Selection of flow metrics for use in hierarchical classification

Flow regime classifications are often based on several flow metrics that together characterize key flow attributes: typically magnitude, frequency, duration, timing, and rate-of-change of streamflow [6, 37, 39]. Dhungel et al. [37] used principal components analysis to identify 16 candidate flow variables from 65 metrics that reflected broadly different dimensions of the five flow attributes. We used identical methods to calculate site-specific values for the same 16 metrics, but we considered 6 other flow metrics related to zero-flow days instead of the two metrics used by Dhungel et al. [37] because we wanted to provide greater resolution in distinguishing different types of nonperennial streamflow patterns. We examined correlations between these 20 candidate metrics and dropped one of any pair of metrics that were correlated (Pearson $r > |0.75|$) with one another. When selecting which metric to use from a pair of correlated metrics, we used the one that we considered most likely have an interpretable effect on stream aquatic life [5, 6, 12, 16, 24]. We ultimately selected 12 metrics for use in the hierarchical cluster analysis (Table 2). The final 12 metrics included 9 used by Dhungel et al. [37] and 3 metrics characterizing different aspects of zero flow conditions: the mean number of zero-flow days per year and its coefficient of variation (1. ZcntMn, 2. ZcntMnCV) and the mean duration of continuous zero-flow days across the period of record (7. Meanzd_R)

Table 2. The 12 flow metrics used in the hierarchical cluster analysis.

Flow metric description	**Abbr.**
1. Mean number of zero-flow days per year	ZcntMn
2. Coefficient of variation of the mean number of zero-flow days per year	ZcntMnCV
3. Seven day moving average of minimum flow	Q7min
4. Bank full flow	BFF
5. Flood duration (mean number of days exceeding bank full flow)	FD
6. Mean days to annual peak flow	Pk_time
7. Mean duration of all zero-flow events across the record	Meanzd_R
8. Mean days to 50% of total annual flow	T50
9. Mean number of low-flow events per year	LFE
10. Mean number of high flow events per year	HFE
11. Constancy: a unitless measure of uncertainty**,** higher C = high certainty throughout year	C
12. Contingency: a unitless measure of seasonal uncertaint**y,** higher M = high certainty by season	M

2.6 Classifying streamflow regimes

Hierarchical cluster analysis has been frequently used to characterize streamflow regimes [37, 38, 42]. For our analyses, we used Ward's method of agglomerative hierarchical cluster analysis [54]. This method produces a dendrogram that describes similarities among sites based on the joint variation among sites in the metrics used for classification. The dendrogram can then be visually inspected to identify increasingly resolved and subtle differences among classes as the number of branches increases [37, 38, 42]. We initially examined up to 8 classes but quickly disregarded one 15-stream subclass because preliminary analyses revealed it was predicted very poorly (>80% class prediction error). We thus explored the remaining 7 classes, but we were primarily interested in how well 3–5 classes partitioned variation in flow regimes. We reasoned that 3 to 5 classes probably represent a trade-off between resolution, which will affect the accuracy and precision of assessments, the ability to predict class membership of ungauged reaches, and the practical needs of water resource managers for a classification scheme that can be easily communicated to stakeholders.

We also created 4 alternate classifications based on the percent of zero-flow values in the stream flow records for comparison with the cluster-based approaches (Table 1). This threshold approach has been previously used to classify nonperennial streamflow regimes in arid regions [25]. For this approach, we calculated both percent of ZFD across the record for each stream as well as the percent of years with at least one zero-flow day, ZFY, across the record for each stream. We then applied different thresholds to define three or four different classes for both the ZFD and ZFY classifications. These thresholds were selected to create classes that generally aligned with the flow characterization methods being considered to support of bioassessments of southwestern USA streams [15, 25], which we also thought could be 1) biologically relevant and 2) predictable with watershed attributes. ZFD thresholds were set at 0% (the 120 perennial streams), >0%, >2%, and >20% ZFD for a four-group classification and at 0%, >0%, and >20% for a three-group ZFD classification. Thresholds for the ZFY classifications were set to create two 3-group classifications with class thresholds of 0% ZFY (the same 120 perennial streams), >0%, and >20% ZFY and another with 0% (the same 120 perennial streams), >0%, and >75% ZFY thresholds.

We also used RF regression models to assess if continuous variation in percent ZFY and percent ZFD was predictably associated with variation in physio-climatic conditions across the study region. The rationale for these analyses was that if the models accounted for much of the variation in ZFY and ZFD, managers could use them to map flow regimes based on any threshold of ZFY or ZFD that was appropriate to their specific management needs.

2.7 Selection of predictor variables

We considered over 500 landscape and climate attributes as candidate predictor variables of streamflow regime classes. Seventy-one of these attributes were obtained from the USEPA's *StreamCat* dataset [59], which includes geographic descriptors such as latitude, longitude, and watershed area as well as watershed-level attributes describing climate, vegetation, soil, and geomorphology, many of which are known to be associated with flow generation and variation in the magnitude, frequency, duration, timing, or rate of change of streamflow [6, 37-42]. For instance, *StreamCat* provides a topographic wetness index and soil wetness index, both of which have been used as predictors of runoff in hydrologic models [60]. We expected these watershed-level attributes to be important predictors of flow regime classes, particularly the base flow index (BFI) [14]. The *StreamCat* BFI was generated through spatial interpolation of USGS gauge-specific BFI calculations, where BFI represents the baseflow volume / total streamflow volume as calculated through a combination of a moving minimum flow, similar to the Q7min (Table 2), and a recession slope test [61, 62].

Given the driving role of evapotranspiration and climate on streamflow patterns in arid landscapes [20, 34, 41, 63], several additional evapotranspiration and aridity indices not available in *StreamCat* were calculated from 30-year, 4-km (1981-2010) mean air temperature and precipitation PRISM data [64] and the modeled runoff index in *StreamCat* [59]. These variables were combined through simple algebra [65]:

Longterm Evapotranspiration (mm) calculated from PRISM precipitation (Precip8110Ws) [59, 64] and *StreamCat* Runoff (RunoffWs) [59] as:

$$ET.mm = Precip8110Ws - RunoffWs,$$

Regional Potential Evapotranspiration (PET) calculated from PRISM temperature (Tmean8110Ws) [59, 64] as:

$$PET = (1.2*10^{10})*exp(-4620/(Tmean8110Ws +273)), \text{ and}$$

An Aridity Index (PET_P) and Evaporative Index (ET_P) calculated as:

$$PET\_P = PET / Precip8110Ws \text{ and}$$

$$ET\_P = ET.mm / Precip8110Ws.$$

We compiled additional watershed-level summaries of several climate variables [47, 63, 66] including the average watershed snow-cover from the Modis-10A1 V6 Snow Cover Daily Global 500m product [67] and a county level drought-severity index from the United States Drought Monitor [66].

We used *ArcGIS* [68] to generate curvature variables from a 30m Digital Elevation Model (DEM) downloaded through Earth Explorer [69]. The *ArcGIS* curvature tool was used to calculate the second derivative of the DEM to generate planar, profile, and combined curvature calculations such as maximum, minimum, mean, and standard deviation. Slope curvature, describing the convexity or concavity of the terrain, can affect runoff [32, 60] through groundwater access [70] and saturation rate of soil profiles in response to rainfall [31, 36, 71]. The summary metrics of the perpendicular and parallel terrain curvature were used as additional candidate predictors.

We selected 95 of the >500 variables as candidate predictors (Appendix A). Predictors were eliminated if they lacked evidence of being associated with either runoff generation, base-flow contribution, or zero-flow events. We also removed predictors that were functionally redundant but calculated differently. In general, we retained the predictors that were most easily calculated to facilitate repeatability in future analyses. We also removed predictors that were missing from one or more of the 287 gaged locations or were constant across the study region.

We then used the *VSURF* function in R [54, 72] to identify the most parsimonious set of predictors that produced the best model performance. *VSURF* uses RF to apply a 3-step variable selection process to further eliminate redundancy and identify the subset of predictor variables that produce the lowest out-of-bag errors for each model. VSURF will optimize variable selection based on either interpretation or prediction, and we chose to optimize predictive accuracy.

2.8 Predicting streamflow regime classes

We used RF models to predict class membership because the RF algorithm typically performs better than many alternative types of classifiers [54, 55]. RF models are especially useful when there are many predictor variables. RF models are also resistant to overfitting and automatically incorporate interactions between predictor variables. RF models create hundreds of classification (or regression) trees based on bootstrapped subsamples of the data and assess model performance based on the overall out-of-bag (OOB, observations withheld when building each tree) prediction error. For each classification model, we set a random sampling limit within each class according to the smallest class size by applying the ‘sampsize’ and ‘strata’ arguments in the RF algorithm. The use of these arguments balanced the class sizes to avoid inherent biases in over-predicting classes with large numbers of observations [73]. We used confusion matrices to evaluate the prediction errors associated with each individual class in each classification. We used pseudo-r^2 values to evaluate the amount of variance RF regression models explained in ZFY and ZFD values.

RESULTS

3.1 Hierarchical classifications

We identified a total of five hierarchical levels of flow regimes from the hierarchical cluster analysis (Figure 3). The first split separated a group of nonperennial streams (Class A) from a group that included both perennial and nonperennial streams (Class B). The second split separated Class A into two subgroups (A1 and A2) and retained Class B. The third split created two subclasses of Class B (B1 and B2), both of which included a mix of perennial and nonperennial streams. The fourth split created two subclasses of B1 (B1a and B1b) resulting in five total classes. B1a included just nonperennial streams, whereas B1b included just perennial streams. The next two splits created two subclasses of both B1b (B1b.i and B1b.ii) and B2 (B2a and B2b) for a total of 7 classes. This 7-group classification included four classes that were wholly or largely composed of nonperennial streams (A1, A2, B1a, and B2a) and three classes that consisted of just perennial streams (B1b.i, B1b.ii, and B2b).

Classes A streams included the driest streams in the dataset and exhibited intermittent flood peaks associated with monsoonal rainfall (Figures 4 and 5). This class had the absolute highest mean annual count of zero-flow days (*ZcntMn*, Figure 4), which approached 300 days for A1 streams and ~150 days for A2 streams. In comparison, Class B streams had ~75 days of zero-flow days. Class A streams ranged from being largely ephemeral with over 75% zero flow and no baseflow (A1) to streams exhibiting monsoonal flow signatures (A2) with between 20 and 75% zero-flow days and some baseflow (Figure 5). For both A1 and A2 subclasses, flood peaks were orders of magnitude higher than in other classes, and baseflows were at or near zero. Both A1 and

A2 subclasses were also nearly identical in the distribution of 7-day moving minimum flows (*Q7min*), and most streams had *Q7min* values several orders of magnitude lower than Class B streams (Figure 4). Class A streams were also characterized by a slightly

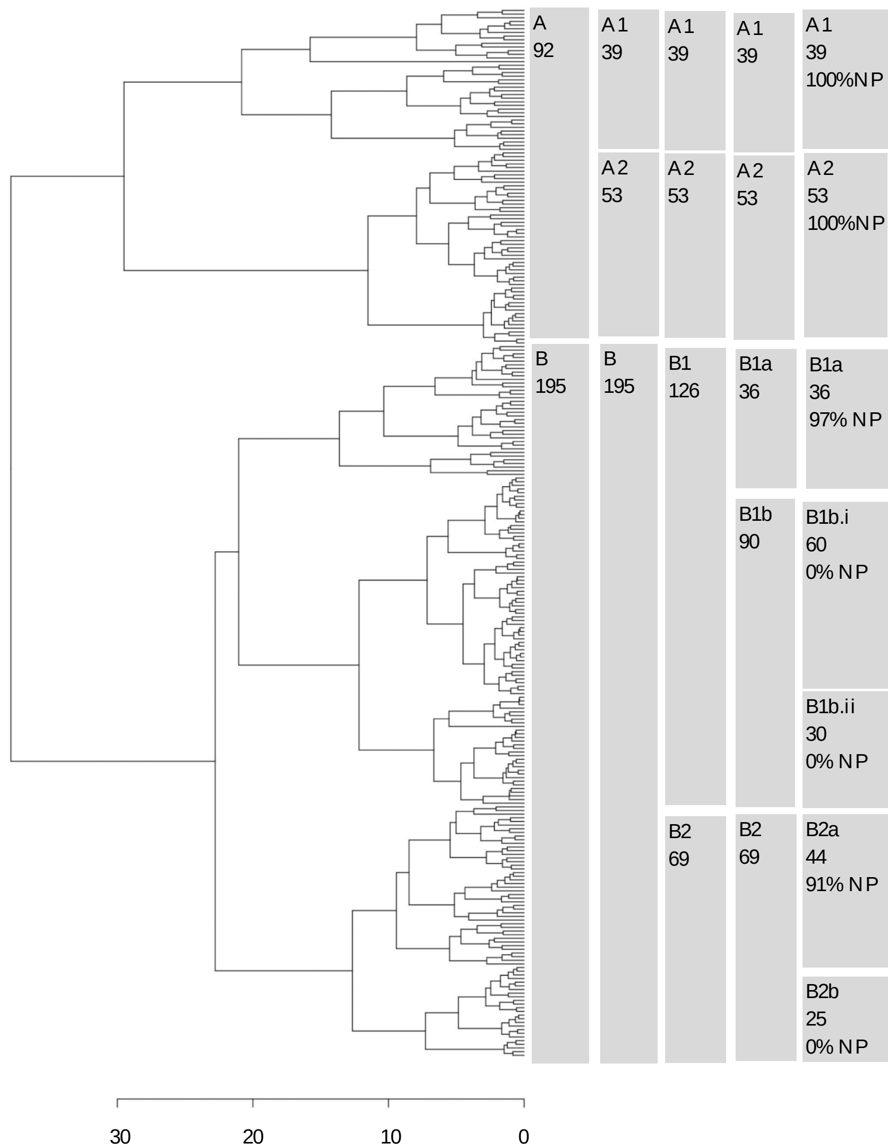

Figure 3. The dendrogram produced by the hierarchical cluster analysis with the different classification levels shown to the right of the dendrogram along with the number of streams in each class and the percent of streams that were nonperennial (NP) in each class.

greater number of both high- and low-flow events (*HFE* and *LFE*) than Class B streams (Figure 4). Most Class A1 streams with over 75% zero-flow days lacked any seasonal pattern, as quantified by the high constancy (*C*) and low contingency (*M*) of flow in these streams (Figure 5). Collectively these metrics describe streams with more irregular flows than Class B streams (Figures 4 and 5).

Class B streams included a mix of perennial and nonperennial streams (Figure 3) characterized by seasonal variations in flow (Figures 4 and 5). Subclasses of B differed in the specific patterns of seasonality they exhibited, which included patterns associated with snowmelt (B1) and rainfall (B2) (Figure 5). Class B1a streams contained 97% nonperennial streams with the absolute highest *ZcntMnCV* (Figure 4), and some of the lowest mean zero-flow day durations across the study streams *(Meanzd_R*, Figure 4). The two subclasses of snowmelt driven B1 streams differed from one another in terms of the presence of zero-flow days - B1a streams were nonperennial and B1b streams were perennial (Figures 3, 4, 5). The subclasses of the rainfall driven B2 streams (Figure 5) were similarly distinguished by one class being nonperennial (B2a) and the other being perennial (B2b).

3.2 Percent zero-flow classifications

Both ZFY classifications included the same class of 120 perennial streams but differed in the number and size of nonperennial classes that were assigned. One ZFY classification included 41 streams ranging from >0 to 20% ZFYs and 126 streams with >20% ZFY. The other ZFY classification included 83 streams ranging from >0 to 75% ZFY and 84 streams with >75% ZFY. The ZFY and ZFD classes were generally similar in the distribution of streams among classes.

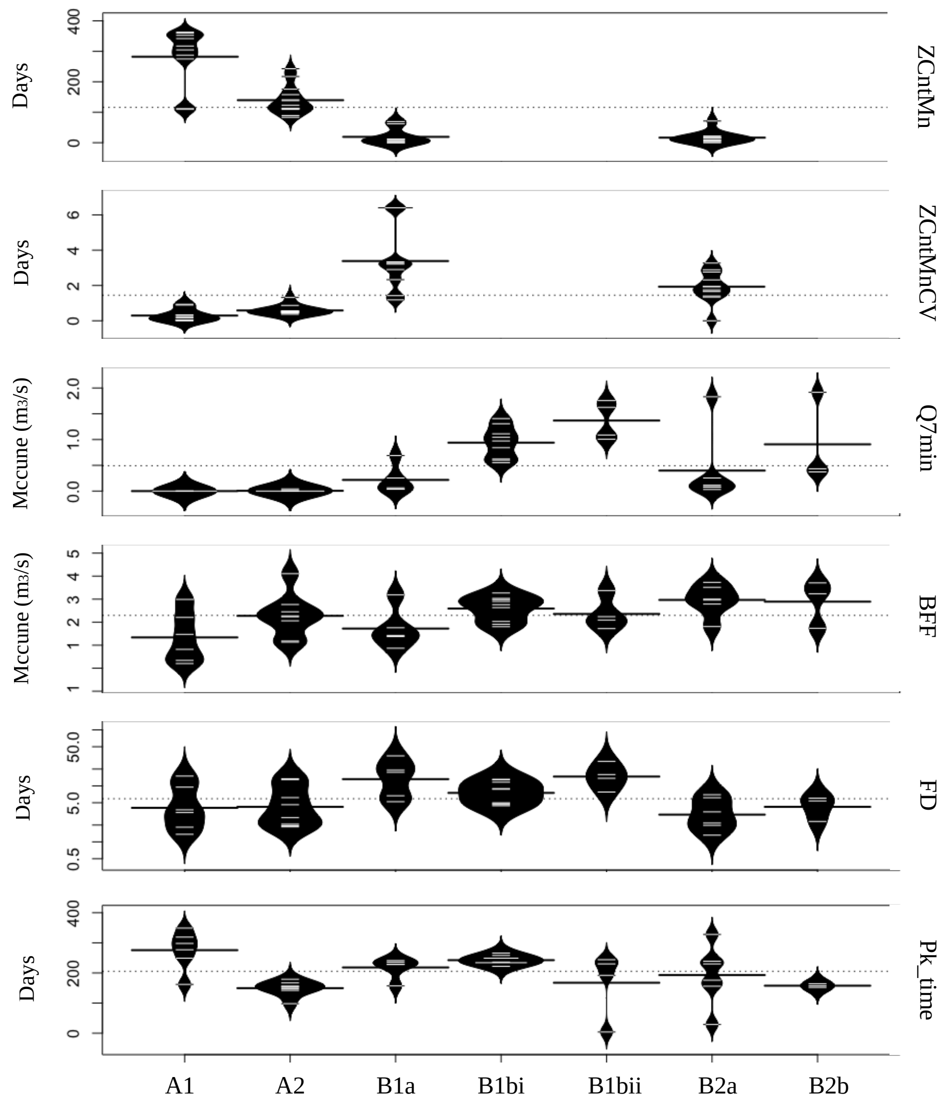

Figure 4a. Density distribution plots of values for the metrics for each of the seven most resolved classes (A1 to B2b). Plots for the different metrics are presented in the same order as given in Table 2.

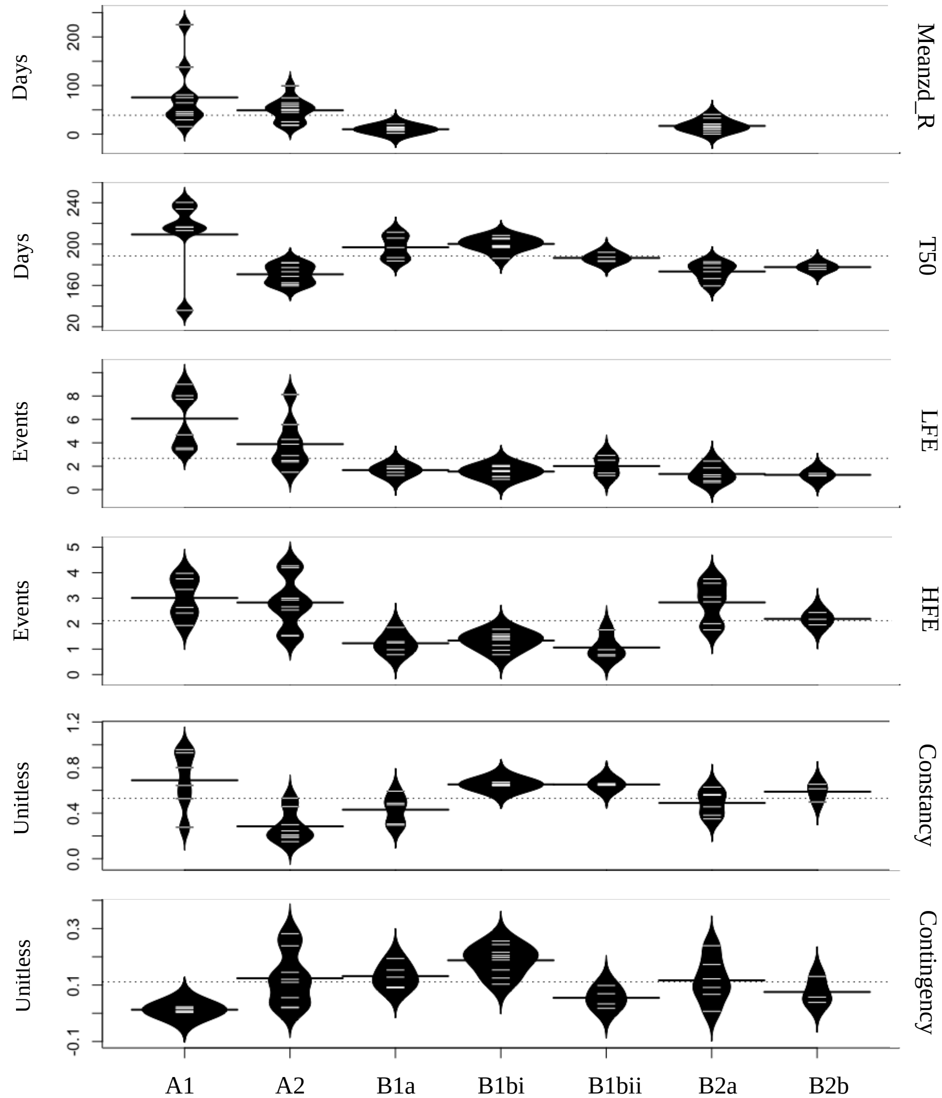

Figure 4b. Density distribution plots of values for the metrics for each of the seven most resolved classes (A1 to B2b). Plots for the different metrics are presented in the same order as given in Table 2.

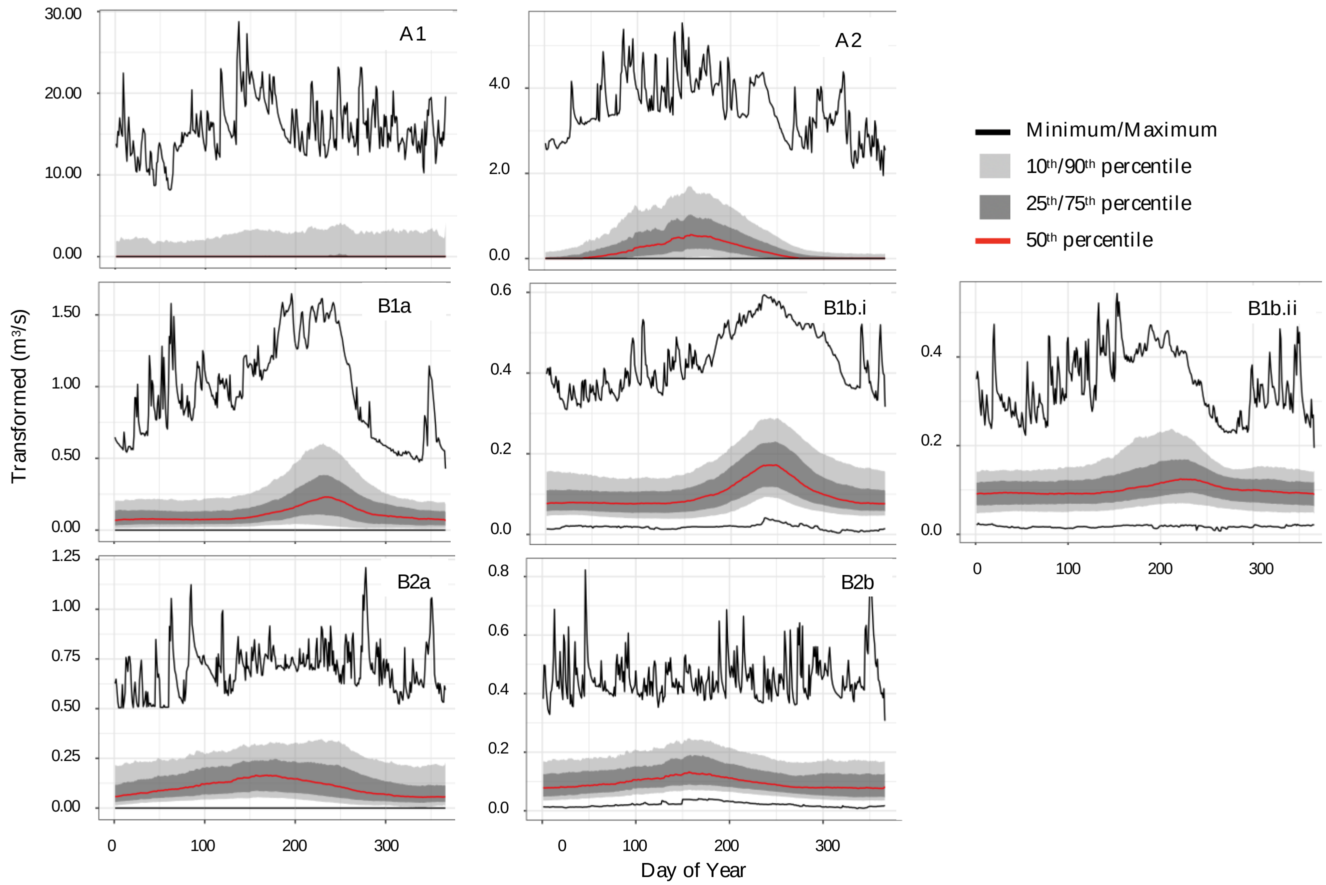

Figure 5. Dimensionless reference hydrographs for the seven most resolved classes (A1 to B2b).

Both ZFD classifications included a class consisting of the 120 perennial streams but differed in the number and size of nonperennial classes. For the four-group ZFD classification, one group included 37 streams with >0 to 2% ZFD, which were streams predominantly from group B1a. Another class consisted of 43 streams with >2 to 20% ZFD, which contained a mix of streams from A2, B1a, and predominantly B2a. The alternative three-group ZFD classification combined these 37 and 43 (80) streams into the >0 to 20% ZFD class. The last class for both ZFD classifications consisted of 87 streams with >20% ZFD, which was almost exclusively class A streams.

3.3 Assessing model performance

Overall OOB prediction error for the RF classification models varied from 15 to 41%, and pseudo-r^2 values varied from 50-58% for the 2 RF regression models (Tables 3 and 4). In general, and as expected, overall prediction error increased as the number of classes increased. Furthermore, class-specific prediction errors sometimes varied markedly – i.e., some types of flow regimes were more difficult to predict than other types.

3.4 Predicting hierarchically-defined classifications

Performance of the models predicting the hierarchical-based stream classes varied both with the number of classes and among classes (Table 3). These models had overall OOB errors of 15, 23, 24, 34, and 39% for the two- through seven-group classifications, respectively. The variation between class-specific prediction errors became more pronounced as the number of groups increased. There were no obvious trends in what types of classes were predicted with greater or lesser accuracy.

Table 3. Confusion matrices for the 2- to 7-group hierarchical classification models, with model performance range of 15% to 39% out-of-bag (OOB) error. Classes with over 90% nonperennial streams are marked with a *. Classes with 100% perennial streams are marked with a †. Observed classes are rows, and predicted classes are columns.

Two classes (15% OOB)

	A*	B	Error (%)	Totals
A*	79	12	13	92
B	32	163	16	195

Three classes (23% OOB)

	A1*	A2*	B	Error	Totals
A1*	32	4	2	16	39
A2*	8	39	6	26	53
B	14	33	148	24	195

Four classes (24% OOB)

	A1*	A2*	B1	B2	Error	Totals
A1*	29	4	2	3	24	39
A2*	7	36	2	8	32	53
B1	2	5	108	11	14	126
B2	7	13	5	44	36	69

Five classes (34% OOB)

	A1*	A2*	B1a*	B1b†	B2	Error	Totals
A1*	28	4	2	1	3	26	39
A2*	5	33	3	2	10	38	53
B1a*	1	4	19	11	1	47	36
B1b†	1	5	13	68	3	24	90
B2	6	16	1	4	42	39	69

Seven classes (39% OOB)

	A1*	A2*	B1a*	B1bi†	B1bii†	B2a*	B2b†	Error	Totals
A1*	31	2	2	0	1	2	0	18	39
A2*	6	34	1	1	5	2	4	36	53
B1a*	1	4	17	6	6	0	2	53	36
B1b.i†	0	1	6	51	1	1	0	15	60
B1b.ii†	1	2	2	7	15	0	3	50	30
B2a*	5	12	0	3	3	13	8	70	44
B2b†	0	4	0	0	2	5	14	44	25

Table 4. Confusion matrix for 4 nonperennial classification models. ET0 = perennial streams (equal to 0% ZFY or ZFD); G0L20y = >0, <20% (greater than zero, less than 20 percent ZFY); G20y = >20% (greater than 20 percent ZFY); G0L75y = >0, <75% (greater than zero, less than 75 percent ZFY); G75y = >75% (greater than 75 percent ZFY); G0LT20 = >0, <20% (greater than zero, less than 20 percent ZFD); G20 = >20% (greater than 20 percent ZFD); G0LT2 = >0, <2% (greater than zero, less than 2 percent ZFD); G2LT20 = >2, <20% (greater than two, less than 20 percent ZFD). Nonperennial ZFY classifications were based on a threshold first set at 20% ZFY and next at 75% ZFY for each of 287 streams. Nonperennial ZFD classifications were based on a threshold first set at 20% ZFD and next at two thresholds: 2% & 20% ZFD. Model performance ranged from 24 to 37% OOB error. Observed classes are rows, and predicted classes are columns.

Zero-flow year (ZFY) thresholds						
20% ZFY (31% OOB)						
	ET0	**G0L20y**		**G20y**	**Error (%)**	**Totals**
ET0	88	20		12	27	126
G0L20y	16	17		8	59	41
G20y	16	16		93	26	87
75% ZFY (30% OOB)					**Class**	**Class**
	ET0	**G0L75y**		**G75y**	**Error**	**Totals**
ET0	90	19		11	25	120
G0L75y	25	44		14	47	83
G75y	6	11		66	20	84
Zero-flow days (ZFD) Thresholds						
20% ZFD (34% OOB)					**Class**	**Class**
	ET0	**G0LT20**		**G20**	**Error**	**Totals**
ET0	92	18		10	23	120
G0LT20	24	32		24	60	80
G20	8	12		66	23	87
2% & 20% ZFD (41% OOB)					**Class**	**Class**
	ET0	**G0LT2**	**G2LT20**	**G20**	**Error**	**Totals**
ET0	80	20	11	9	33	120
G0LT2	16	10	5	6	73	37
G2LT20	8	6	18	11	58	43
G20	7	3	15	61	29	87

3.5 Predicting ZFD and ZFY classes and continuous variation

Overall model performance also varied across then ZFY and ZFD classifications (30 – 41% error) (Table 4). In general, the ZFY and ZFD models could not predict classes with intermediate values of ZFY or ZFD, exceeding 70% and sometimes 80% class prediction error.

The results of the ZFY and ZFD RF models were consistent with results of the ZFY and ZFD threshold classification models (Figure 6). These two regression models explained slightly more than 50% of the variation in percent ZFY and ZFD. The prediction errors were sometimes extreme. For example, several streams with < 20% ZFYs were predicted as perennial streams, especially streams with low to just one day of zero flow per year of record. The same pattern occurred for ZFD predictions. However, the ability of these models to predict the large portion (120) of perennial streams and also the component of highly nonperennial streams (>20% ZFY or ZFD) resulted in reasonable model results (Figure 6).

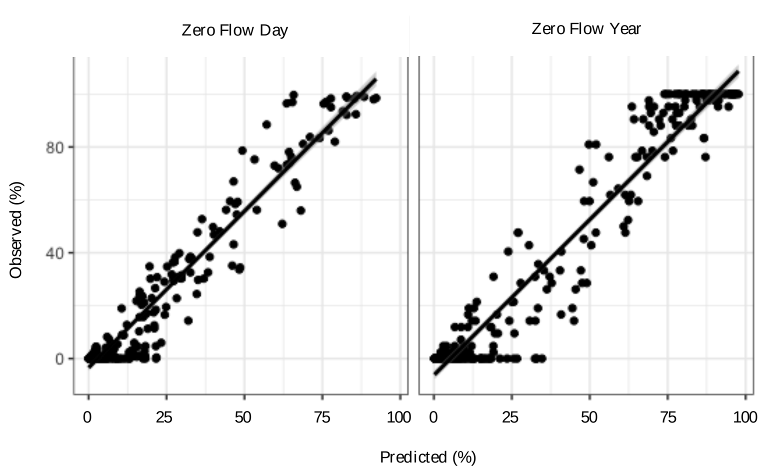

Figure 6. Plots of observed versus predicted values for the ZFY (a) and ZFD (b) regression models.

3.6 Selected predictor variables

A few predictor variables were consistently selected across models Tables S1 and S2. BFI was almost always identified as the top predictor. The only exception was for the hierarchical four-class model that ranked longitude before BFI. The other top predictors were related to location, topography, and climate. Other predictors included the curvature slope of the streambed (typically ParCurveSTD), the PRISM temperatures (typically Tmax8110Ws), evapotranspiration (usually PET), and elevation (usually ElevWs). Overall, some representation of curvature, climate, and location were combined with landcover variables related to snow (MODISsnow) or vegetation (percent grassland, PctGrs2001Ws, and percent forestland, PctMxFst2001Ws) coverage of the watershed (Supplemental Tables S1 and S2). These predictors were also selected for the ZFY and ZFD classification and regression models, although additional dimensions of streambed curvature (MIN_curve, MAX_curve, PerpCurve, etc) and landcover (PctCrop, PctBl, Slp10) were selected for these models as well.

DISCUSSION

Water resource managers currently lack the tools needed to assess the physical, chemical, and biological integrity of the majority of stream ecosystems in arid regions [1]. This problem is especially severe for nonperennial streams, which are understudied relative to perennial streams but which represent the vast majority of the stream networks in these regions [13-15, 20, 21]. Knowledge of how physicochemical factors structure biological communities provides the basis for defining and identifying the reference conditions on which assessments of individual waterbodies are based [3, 11] and is thus central to successful assessment and management programs.

Of the many naturally occurring physicochemical factors that can influence aquatic life in streams, the flow regime is thought to be of paramount importance [5-9]. In arid regions, variation in flow regimes is likely to be especially critical to aquatic life given that a conspicuous dimension of these regimes is the extent to which flows completely cease – a naturally occurring disturbance that has profound effects on the type of biota that can persist in a stream [16-19]. Characterizing, or classifying, natural flow regimes in arid-region streams (and hence their reference conditions) is therefore a critical need in the USA and elsewhere [15, 20-23, 25-31]. Spatially sparse and temporally incomplete flow records typically limit our ability to robustly describe the variability of naturally occurring flow regimes in arid regions [20, 45]. The lack of classification schemes beyond the traditional division of perennial and nonperennial streams has likely limited our appreciation of the full diversity of ways that flow regimes may vary across these landscapes [1, 21]. In this study, we first showed that we could improve the spatial coverage of daily streamflow data by recreating data that were missing from gauges with incomplete records. That process, together with the inclusion of additional flow metrics characterizing different aspects of zero flow, allowed us to explore several alternative classifications that may be of use to water resource managers. Below, we address both the potential utility of these classifications and factors that may limit their use, including our ability to predict where they occur.

4.1 The value of creating standard, long-term synthetic flow records

Developing robust classifications of flow regimes requires data that cover the full range of variability in the flow. For arid regions, this means that flow records should span a full range of nonperennial streams [12-15, 21], and more than one metric related to the frequency, duration, and timing of zero flows are likely needed to describe differences in their flow regimes [21, 22]. By using RF models to estimate the daily flows occurring at gauges with incomplete records from gauges with complete records, we greatly enhanced our ability to characterize the spatial and temporal diversity of streamflows that occur in the southwestern USA [25, 37, 45, 46]. The majority, 55%, of these models explained well over 90% of the variation in daily flows that were recorded in the 197 gauges with partial-records although, on average, flows in nonperennial streams were predicted with slightly less accuracy than those in perennial streams (Table 1). Previously used approaches to classify and predict flow regimes would have disregarded records from up to 145 gauges in our study region because they would not have met data-quality criteria – e.g., they either had less than 10-20 years of record, or they had data gaps greater than 5-10-years [43-45]. Furthermore, a majority of these 145 incomplete records fell within gauge-sparse arid ecoregions in Arizona, New Mexico, and Texas, 50 of which had been excluded by Dhungel et als. [37] in their analysis of the potential effect of climate change on flow regimes across the conterminous USA. By creating synthetic records, we greatly enhanced our ability to characterize variation across the region in all aspects of flow, especially those associated with zero flow (ZFY, ZFD).

4.2 Beyond the perennial-nonperennial dichotomy

Water resource managers in the southwestern USA are starting to realize that a simple dichotomy of perennial and nonperennial streams is unlikely to support management goals [e.g., 15, 25], but certain types of nonperennial streams should be considered separately from other streams. The dendrogram immediately clustered (1^{st}-level split, A and B, Figure 3) one wholly nonperennial group (A), containing $1/3^{rd}$ of the 287 streams. Furthermore, the 2^{nd}-level (three classes) split in the dendrogram defined two nonperennial classes (A1 and A2, Figure 3). These two A classes are similar to two classes identified by McManamay and DeRolph [40], who developed a flow regime classification that spanned the entire conterminous USA. The distinction between these two classes may be critical from a management perspective because states like Arizona must identify ephemeral streams from other nonperennial streams for regulatory purposes. It is unclear, however, how tightly our ephemeral class (A1) matches definitions of ephemeral used by different states. These details notwithstanding, the classification of different types of nonperennial streams as shown here is a critical step in the development and application of consistent, statewide regulatory policies designed to protect regulated waterbodies.

The idea to consider nonperennial streams as distinct from all perennial streams may be over-simplistic. From the hierarchical analysis, pure perennial stream classes were unresolved until the 4^{th}-level (five classes) of the dendrogram – e.g., one of five classes (B1b). A reasonable dichotomy required several hierarchical levels (5) and highly refined classes (5^{th} level, 7-classes, Table 3), where three of seven classes (B1b.i, B1b.ii, B2a) were perennial and four of seven classes (A1, A2, B1a, B2b) contained 90-100% nonperennial streams (Figure 3). However, the last classes formed, nonperennial B2a and perennial B2b, had unreasonable prediction errors. These two classes had highly similar flow regimes (Figure 4) with the driving difference being the presence (B2a) and absence (B2b) of zero flow. Each of these classes is likely to be misclassified as its counter-class, e.g., a B2a stream misclassified as a B2b stream (Table 3). The similarities in flow regimes could result in either high rates of misclassification, and/or undetectable biological differences [18], which would be essential to the design and calibration of some bioassessment tools [74].

We must also emphasize that the classifications that emerged from our analyses are dependent on the specific flow metrics we used to create classes. The metrics used in the hierarchical classification and the thresholds used to define ZFD and ZFY classes were guided by our general ecological knowledge of the importance of flow regimes to aquatic life [16, 37]. For example, zero-flow events represent severe disturbances that can reset the communities of stream ecosystems [17-19, 24], and the number, timing, duration of such events should influence the magnitude and predictability of the recovery dynamics of stream communities [21, 35]. However, until these classifications are tested to assess whether they are actually associated with variation in valued ecological attributes, we cannot guarantee that the types of flow regimes we identified will be useful or how many classes will be needed. Managers ultimately need classifications that allow enough partitioning of biological variation to produce indices that are precise enough to detect ecologically important impairment but do not have so many classes that they are difficult to communicate and unwieldy to use. Assessing the strength of these relationships is an active area of research. Use of a different set of flow metrics would likely lead to different types and numbers of classes [9], and it is possible that such alternative classifications might be more useful than the ones developed here.

4.3 Congruence between hierarchical and zero-flow day classifications

We have focused on the extent to which classes derived from the cluster analysis were interpretable and potentially of regulatory and ecological relevance. However, we note that there was some congruence between classes based on zero-flow days and those based on hierarchical classifications. This congruence was not unexpected given the importance of zero-flow metrics in defining the hierarchical classes. The four most resolved nonperennial classes progressively increased in zero-flow days in this order: B1a, B2b, A2, A1. The B classes were characterized by a lower mean number of zero-flow days per year (50-75 days *ZcntMn,* Table 2) than the A classes (~150 days *ZcntMn* for A2 and ~300 days for A1, Figure 3a). B1a and B2b streams tend to align with nonperennial streams with 0-20% ZFY and ZFD, and A2 and A1 streams tend to align with nonperennial streams with over 75% ZFY and over 20% ZFD (Table 4). B1a streams are on the lower end of 0-20% ZFY/ZFD range because of irregular years of zero-flow occurrence, whereas the A1 class aligns well with the highly nonperennial >20% ZFY and ZFD classes, which were relatively well predicted (Table 4).

4.4 Predictive variables and prediction errors

Ideally, predictive models will be both accurate (low error) and interpretable in terms of what physicoclimatic factors are most important in producing different flow regimes. In this study, both overall and class-specific prediction error rates varied considerably. We observed an expected tradeoff between number of classes and prediction error, which can be used to inform recommendations regarding the practical use of specific classification schemes. If we assume that 25% error is tolerable to water resource managers, our results imply that no more than four metric-based classes will be useful to managers and that none of the zero-flow based classifications would be acceptable. Moreover, in the four-group metric-based classification, the fact that only two classes were predicted with <25% error implies that classification may be of limited utility. These error rates do not seem to be unique to our study. For example, the errors that Dhungel et al [37] observed in predicting eight classes of flow regime ranged from 14 to 43%, with the most error in small nonperennial streams and the least error in small perennial streams. In contrast, McManamay and DeRolph [40] report overall prediction errors of ~5 to 34% for classifications based on 2 to 30 hierarchically-defined flow regime classes across the entire conterminous USA. They did not report class-specific error rates. Their analyses did include three classes of streams for the 30-group classification described as intermittent flashy, but only two stream classes appear to exist in the southwestern USA. We expect the higher error rates that we observed in our study occurred because our classification of southwestern streams was much more resolved due to the increased coverage and length of nonperennial stream gauge records. It appears that we can predict the ends of the ephemeral to perennial continuum reasonably well, but our models have difficulty distinguishing streams that are more variable in the number of zero-flow days from year to year.

Empirical models can perform poorly if predictor variables that represent important drivers of streamflow patterns are missing from the models. We attempted to ensure that the predictor variables we used represented the range of physico-climatic controls on flow regimes expected to occur in arid regions. The specific variables that the VSURF procedure identified as optimal sets of predictors in the models were generally similar across classifications and interpretable in terms of their likely mechanisms. For example, BFI, which represents the slowly varying portion of reliable streamflow [32, 75, 76], was consistently ranked as a top predictor (Tables S1-S3), likely due to its clear linkage to the streamflow patterns themselves. However, it may be unlikely that prediction errors can be substantially improved by incorporating additional predictor variables, and we suspect that flow regimes of individual streams in arid regions may be inherently difficult to predict because physico-climatic conditions and streamflow generation are not as tightly coupled in arid regions as they are in more mesic regions. The physical controls on streamflow response throughout arid regions are typically related to intermittent monsoonal climate patterns, low annual precipitation [34, 37], comparatively high annual rates of potential evapotranspiration [41, 77], and shallow bedrock, leading to lower annual baseflows and often irregular patterns of zero flow [31, 35, 36]. As a result, the arid region rainfall and runoff are often poorly correlated and have non-linear dependency on antecedent conditions [20, 78]. Aryal et al. [51] found that, similar to other studies in semi-arid regions, streamflow patterns were much more muted than precipitation patterns, the spatial distribution of streamflow did not correlate with the spatial pattern of precipitation, and there was no significant correlation between the annual number of no flow days and days with precipitation, or between catchment area and mean annual or seasonal runoff.

Ultimately, any useful classification of streamflow regimes must either generate new understanding of the processes that influence streamflow or help managers identify streams that differ with respect to their management objectives. Our study focused on developing classifications of stream-flow regimes that would allow resource managers to more fully characterize the diversity of flow regimes that occur in arid landscapes. Classifications that extend beyond the traditional dichotomy of perennial and nonperennial streams are almost certainly needed to support developing programs tasked with assessing whether streams are meeting physical, chemical, and biological water quality standards. At a minimum, our study showed that it is possible to distinguish a subset of nonperennial streams that are ephemeral and hence not subject to the environmental protections often afforded to other streams in the USA. An equally important goal is to better understand what level of resolution in flow-regime classifications is needed to best support developing bioassessment programs. Our study adds to a growing body of knowledge that is needed to address that question for stream ecosystems in the arid western USA.

REFERENCES

1. Chessman, B. C.; Thurtell, L. A.; Royal, M. J. Bioassessment in a Harsh Environment: A Comparison of Macroinvertebrate Assemblages at Reference and Assessment Sites in an Australian Inland River System. *Environ. Monit. Assess.* **2006**, *119* (1-3), 303–330. https://doi.org/10.1007/s10661-005-9027-2.

2 .Barbour, M. T.; Gerritsen, J.; Snyder, B. D.; Stribling, J. B.; Others. *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish*; US Environmental Protection Agency, Office of Water Washington, DC, **1999**, Vol. 339. https://gis.lic.wisc.edu/wwwlicgf/glifwc/PolyMet/SDEIS/references/USEPA%202012b.pdf

3. Hawkins, C. P.; Olson, J. R.; Hill, R. A. The Reference Condition: Predicting Benchmarks for Ecological and Water-Quality Assessments. *J. North Am. Benthol. Soc.* **2010**, *29* (1), 312–343. https://doi.org/10.1899/09-092.1.

4. Stubbington, R.; Chadd, R.; Cid, N.; Csabai, Z.; Miliša, M.; Morais, M.; Munné, A.; Pařil, P.; Pešić, V.; Tziortzis, I.; et al. Biomonitoring of Intermittent Rivers and Ephemeral Streams in Europe: Current Practice and Priorities to Enhance Ecological Status Assessments. *Sci. Total Environ.* **2018**, *618*, 1096–1113. https://doi.org/10.1016/j.scitotenv.2017.09.137.

5. Chinnayakanahalli, K. J.; Hawkins, C. P.; Tarboton, D. G.; Hill, R. A. Natural Flow Regime, Temperature and the Composition and Richness of Invertebrate Assemblages in Streams of the Western United States. *Freshw. Biol.* **2011**, *56* (7), 1248–1265.Abatzoglou, J. T.; Barbero, R.; Wolf, J. W.; Holden, Z. A. Tracking Interannual Streamflow Variability with Drought Indices in the U.S. Pacific Northwest. *J. Hydrometeorol.* **2014**, *15* (5), 1900–1912. https://doi.org/10.1175/JHM-D-13-0167.1.

6. Poff, N. L.; Allan, J. D.; Bain, M. B.; Karr, J. R.; Prestegaard, K. L.; Richter, B. D.; Sparks, R. E.; Stromberg, J. C. The Natural Flow Regime. *Bioscience* **1997**, *47* (11), 769–784. https://doi.org/10.2307/1313099.

7. Poff, N. L.; Ward, J. V. Implications of Streamflow Variability and Predictability for Lotic Community Structure: A Regional Analysis of Streamflow Patterns. *Can. J. Fish. Aquat. Sci.* **1989**, *46* (10), 1805–1818. https://doi.org/10.1139/f89-228.

8. Poff, N. L.; Zimmerman, J. K. H. Ecological Responses to Altered Flow Regimes: A Literature Review to Inform the Science and Management of Environmental Flows. *Freshw. Biol.* **2010**, *55* (1), 194–205. https://doi.org/10.1111/j.1365-2427.2009.02272.x.

9. Yarnell, S; Stein, E.; Webb, J.; Grantham, T.; Lusardi, R.; Zimmerman, J.; Peek, R.; Lane, B.; Howard, J.; Sandoval-Solis, S. A functional flows approach to selecting ecologically relevant flow metrics for environmental flow applications. *River Research and Applications* 2020, 36(2): 318-24. https://doi.org/10.1002/rra.3575.

10. Ode, P. R.; Rehn, A. C.; Mazor, R. D.; Schiff, K. C.; Stein, E. D.; May, J. T.; Brown, L. R.; Herbst, D. B.; Gillett, D.; Lunde, K.; et al. Evaluating the Adequacy of a Reference-Site Pool for Ecological Assessments in Environmentally Complex Regions. *Freshw. Sci.* **2016**. https://doi.org/10.1086/684003.

11. Stoddard, J. L.; Larsen, D. P.; Hawkins, C. P.; Johnson, R. K.; Norris, R. H. Setting Expectations for the Ecological Condition of Streams: The Concept of Reference Condition. *Ecol. Appl.* **2006**, *16* (4), 1267–1276. https://doi.org/10.1890/1051-0761(2006)016.

12. Datry, T.; Larned, S. T.; Fritz, K. M.; Bogan, M. T.; Wood, P. J.; Meyer, E. I.; Santos, A. N. Broad-Scale Patterns of Invertebrate Richness and Community Composition in Temporary Rivers: Effects of Flow Intermittence. *Ecography* 2014, *37* (1), 94–104. https://doi.org/10.1111/j.1600-0587.2013.00287.x.

13. Busch, M. H.; Costigan, K. H.; Fritz, K. M.; Datry, T.; Krabbenhoft, C. A.; Hammond, J. C.; Zimmer, M.; Olden, J. D.; Burrows, R. M.; Dodds, W. K.; Others. What’s in a Name? Patterns, Trends, and Suggestions for Defining Non-Perennial Rivers and Streams. *Water* 2020, *12* (7), 1980. https://doi.org/10.3390/w12071980.

14. Datry, T.; Bonada, N.; Boulton, A. J. Chapter 6 - Conclusions: Recent Advances and Future Prospects in the Ecology and Management of Intermittent Rivers and Ephemeral Streams. In *Intermittent Rivers and Ephemeral Streams*; Datry, T., Bonada, N., Boulton, A., Eds.; Academic Press, **2017**; pp 563–584. https://doi.org/10.1016/B978-0-12-803835-2.00031-0.

15. Levick, L., J. Fonseca, D. Goodrich, M. Hernandez, D. Semmens, J. Stromberg, R. Leidy, M. Scianni, D. P. Guertin, M. Tluczek, and W. Kepner. *The Ecological and Hydrological Significance of Ephemeral and Intermittent Streams in the Arid and Semi-Arid American Southwest*; EPA/600/R-08/134, ARS/233046; U.S. Environmental Protection Agency and USDA/ARS Southwest Watershed Research Center, **2008**. https://www.epa.gov/sites/production/files/2015-03/documents/ephemeral_streams_report_final_508-kepner.pdf

16. Feminella, J. W. Comparison of Benthic Macroinvertebrate Assemblages in Small Streams along a Gradient of Flow Permanence. *J. North Am. Benthol. Soc.* **1996**, *15* (4), 651–669. https://doi.org/10.2307/1467814.

17. Giam, X.; Chen, W.; Schriever, T. A.; Van Driesche, R.; Muneepeerakul, R.; Lytle, D. A.; Olden, J. D. Hydrology Drives Seasonal Variation in Dryland Stream Macroinvertebrate Communities. *Aquat. Sci.* **2017**, *79* (3), 705–717. https://doi.org/10.1007/s00027-017-0530.

18. Lunde, K. B.; Cover, M. R.; Mazor, R. D.; Sommers, C. A.; Resh, V. H. Identifying Reference Conditions and Quantifying Biological Variability Within Benthic Macroinvertebrate Communities in Perennial and Non-Perennial Northern California Streams. *Environ. Manage.* **2013**, *51* (6), 1262–1273. https://doi.org/10.1007/s00267-013-0057-1.

19. Schriever, T. A.; Bogan, M. T.; Boersma, K. S.; Cañedo-Argüelles, M.; Jaeger, K. L.; Olden, J. D.; Lytle, D. A. Hydrology Shapes Taxonomic and Functional Structure of Desert Stream Invertebrate Communities. *Freshw. Sci.* **2015**, *34* (2), 399–409. https://doi.org/10.1086/680518.

20. Hughes, D. A. Modelling Semi-Arid and Arid Hydrology and Water Resources: The Southern Africa Experience. *Hydrological modelling in arid and semi-arid areas* **2008**, 29–40. https://pdfs.semanticscholar.org/a555/ffaec9c39fdf1cb870601267bf87f80291d9.pdf

21. Kennard, M. J.; Pusey, B. J.; Olden, J. D.; Mackay, S. J.; Stein, J. L.; Marsh, N. Classification of Natural Flow Regimes in Australia to Support Environmental Flow Management. *Freshw. Biol.* **2010**, *55* (1), 171–193. https://doi.org/10.1111/j.1365-2427.2009.02307.x.

22. Tavassoli, H. R.; Tahershamsi, A.; Acreman, M. Classification of Natural Flow Regimes in Iran to Support Environmental Flow Management. *Hydrol. Sci. J.* 2014, *59* (3-4), 517–529. https://doi.org/10.1080/02626667.2014.890285.

23. Larned, S. T.; Datry, T.; Arscott, D. B.; Tockner, K. Emerging Concepts in Temporary-River Ecology. *Freshw. Biol.* 2010, *55* (4), 717–738. https://doi.org/10.1093/biosci/bit027

24. Stubbington, R.; Paillex, A.; England, J.; Barthès, A.; Bouchez, A.; Rimet, F.; Sánchez-Montoya, M. M.; Westwood, C. G.; Datry, T. A Comparison of Biotic Groups as Dry-Phase Indicators of Ecological Quality in Intermittent Rivers and Ephemeral Streams. *Ecol. Indic.* **2019**, *97*, 165–174. https://doi.org/10.1016/j.ecolind.2018.09.061.

25. Anning, D. W.; Parker, J. T. C. Predictive Models of the Hydrological Regime of Unregulated Streams in Arizona. *U.S. Geological Survey Open-File Report.* **2009**, *1269*, 33 pgs. http://pubs.usgs.gov/of/2009/1269/.

26. Belmar, O.; Velasco, J.; Martinez-Capel, F. Hydrological Classification of Natural Flow Regimes to Support Environmental Flow Assessments in Intensively Regulated Mediterranean Rivers, Segura River Basin (Spain). *Environ. Manage.* **2011**, *47* (5), 992–1004. https://doi.org/10.1007/s00267-011-9661-0.

27. Berhanu, B.; Seleshi, Y.; Demisse, S. S.; Melesse, A. M. Flow Regime Classification and Hydrological Characterization: A Case Study of Ethiopian Rivers. *Water* **2015**, *7* (6), 3149–3165. https://doi.org/10.3390/w7063149.

28. Nikolaidis, N. P.; Demetropoulou, L.; Froebrich, J.; Jacobs, C.; Gallart, F.; Prat, N.; Porto, A. L.; Campana, C.; Papadoulakis, V.; Skoulikidis, N.; et al. Towards Sustainable Management of Mediterranean River Basins: Policy Recommendations on Management Aspects of Temporary Streams. *Water Policy* **2013**, *15* (5), 830–849.

29. Skoulikidis, N. T.; Sabater, S.; Datry, T.; Morais, M. M.; Buffagni, A.; Dörflinger, G.; Zogaris, S.; Del Mar Sánchez-Montoya, M.; Bonada, N.; Kalogianni, E.; et al. Non-Perennial Mediterranean Rivers in Europe: Status, Pressures, and Challenges for Research and Management. *Sci. Total Environ.* **2017**, *577*, 1–18. https://doi.org/10.1016/j.scitotenv.2016.10.147.

30. Levick, L.; Hammer, S.; Lyon, R.; Murray, J.; Birtwistle, A.; Guertin, P.; Goodrich, D.; Bledsoe, B.; Laituri, M. An Ecohydrological Stream Type Classification of Intermittent and Ephemeral Streams in the Southwestern United States. *J. Arid Environ.* **2018**, *155*, 16–35. https://doi.org/10.1016/j.jaridenv.2018.01.006.

31. Sutfin, N. A.; Shaw, J. R.; Wohl, E. E.; Cooper, D. J. A Geomorphic Classification of Ephemeral Channels in a Mountainous, Arid Region, Southwestern Arizona, USA. *Geomorphology* **2014**, *221*, 164–175. https://doi.org/10.1016/j.geomorph.2014.06.005.

32. Payn, R. A.; Gooseff, M. N.; McGlynn, B. L.; Bencala, K. E.; Wondzell, S. M. Exploring Changes in the Spatial Distribution of Stream Baseflow Generation during a Seasonal Recession: Exploring Spatial Distribution of Stream Baseflow. *Water Resour. Res.* **2012**, *48* (4), 331. https://doi.org/10.1029/2011WR011552.

33. Goodrich, D. C.; Kepner, W. G.; Levick, L. R.; Wigington, P. J., Jr. Southwestern Intermittent and Ephemeral Stream Connectivity. J. Am. Water Resour. Assoc. 2018, 54 (2), 400–422. https://doi.org/10.1111/1752-1688.12636.

34. Reynolds, L. V.; Shafroth, P. B.; LeRoy Poff, N. Modeled Intermittency Risk for Small Streams in the Upper Colorado River Basin under Climate Change. *J. Hydrol.* **2015**, *523*, 768–780. https://doi.org/10.1016/j.jhydrol.2015.02.025.

35. Costigan, K. H.; Jaeger, K. L.; Goss, C. W.; Fritz, K. M.; Goebel, P. C. Understanding Controls on Flow Permanence in Intermittent Rivers to Aid Ecological Research: Integrating Meteorology, Geology and Land Cover: Integrating Science to Understand Flow Intermittence. *Ecohydrol.* **2016**, *9* (7), 1141–1153. https://doi.org/10.1002/eco.1712.

36. Costigan, K. H.; Kennard, M. J.; Leigh, C.; Sauquet, E.; Datry, T.; Boulton, A. J. Chapter 2.2 - Flow Regimes in Intermittent Rivers and Ephemeral Streams. In *Intermittent Rivers and Ephemeral Streams*; Datry, T., Bonada, N., Boulton, A., Eds.; Academic Press, **2017**; pp 51–78. https://doi.org/10.1016/B978-0-12-803835-2.00003-6.

37. Dhungel, S.; Tarboton, D. G.; Jin, J.; Hawkins, C. P. Potential Effects of Climate Change on Ecologically Relevant Streamflow Regimes. *River Res. Appl.* **2016**, *32* (9), 1827–1840. https://doi.org/10.1002/rra.3029.

38. Lane, B. A.; Dahlke, H. E.; Pasternack, G. B.; Sandoval-Solis, S. Revealing the Diversity of Natural Hydrologic Regimes in California with Relevance for Environmental Flows Applications. *J. Am. Water Resour. Assoc.* **2017**, *53* (2), 411–430. https://doi.org/10.1111/1752-1688.12504.

39. Carlisle, D. M.; Falcone, J.; Wolock, D. M.; Meador, M. R.; Norris, R. H. Predicting the Natural Flow Regime: Models for Assessing Hydrological Alteration in Streams. *River Res. Appl.* **2009**, *8*. https://doi.org/10.1002/rra.1247.

40. McManamay, R. A.; DeRolph, C. R. A Stream Classification System for the Conterminous United States. *Sci Data* **2019**, *6*, 190017. https://doi.org/10.1038/sdata.2019.17.

41. Weiß, M.; Menzel, L. A Global Comparison of Four Potential Evapotranspiration Equations and Their Relevance to Stream Flow Modelling in Semi-Arid Environments. In *Advances and visions in large-scale hydrological modelling - 11th Workshop on Large-scale Hydrological Modelling, Kelkheim-Eppenhain, Germany, 31 October–2 November 2007*; Copernicus GmbH, **2008**; Vol. 18, pp 15–23. https://doi.org/10.5194/adgeo-18-15-2008.

42. Olden, J. D.; Kennard, M. J.; Pusey, B. J. A Framework for Hydrologic Classification with a Review of Methodologies and Applications in Ecohydrology. *Ecohydrol.* **2012**, *5* (4), 503–518. https://doi.org/10.1002/eco.251.

43. Deweber, J. T.; Tsang, Y.-P.; Krueger, D. M.; Whittier, J. B.; Wagner, T.; Infante, D. M.; Whelan, G. Importance of Understanding Landscape Biases in USGS Gage Locations: Implications and Solutions for Managers. *Fisheries* **2014**, *39* (4), 155–163. https://doi.org/10.1080/03632415.2014.891503.

44. Vörösmarty, C. J.; Fekete, B.; Tucker, B. A. River Discharge Database. *The Institute for the Study of Earth, Oceans, and Space, University of New Hampshire,* **1998.** http://www. rivdis. sr. unh. edu .

45. Auerbach, D. A.; Buchanan, B. P.; Alexiades, A. V.; Anderson, E. P.; Encalada, A. C.; Larson, E. I.; McManamay, R. A.; Poe, G. L.; Walter, M. T.; Flecker, A. S. Towards Catchment Classification in Data-Scarce Regions. *Ecohydrol.* **2016**, *9* (7), 1235–1247. https://doi.org/10.1002/eco.1721.

46. Giustarini, L., Parisot, O., Ghoniem, M., Hostache, R., Trebs, I. and Otjacques, B. A User-Driven Case-Based Reasoning Tool for Infilling Missing Values in Daily Mean River Flow Records. *Environmental Modelling & Software*. **2016**, *82*, 308-320. https://doi.org/10.1016/j.envsoft.2016.04.013.

47. Douville, H.; Peings, Y.; Saint-Martin, D. Snow-(N)AO Relationship Revisited over the Whole Twentieth Century. *Geophys. Res. Lett.* **2017**, *44* (1), 569–577. https://doi.org/10.1002/2016GL071584.

48. Jaeger, K. L.; Sando, R.; McShane, R. R.; Dunham, J. B.; Hockman-Wert, D. P.; Kaiser, K. E.; Hafen, K.; Risley, J. C.; Blasch, K. W. Probability of Streamflow Permanence Model (PROSPER): A Spatially Continuous Model of Annual Streamflow Permanence throughout the Pacific Northwest. *Journal of Hydrology X* **2019**, *2*, 100005. https://doi.org/10.1016/j.hydroa.2018.100005.

49. Prancevic, J.; Kirchner, J. W. Topographic Controls on the Expansion and Contraction of Stream Networks; **2018**, Vol. 2018, p H31L – 2112. https://doi.org/10.1029/2018GL081799

50. Ward, A. S.; Schmadel, N. M.; Wondzell, S. M. Simulation of Dynamic Expansion, Contraction, and Connectivity in a Mountain Stream Network. *Adv. Water Resour.* **2018**, *114*, 64–82. https://doi.org/10.1016/j.advwatres.2018.01.018.

51. Aryal, S. K.; Zhang, Y.; Chiew, F. Enhanced Low Flow Prediction for Water and Environmental Management. *J. Hydrol.* **2020**, *584*, 124658. https://doi.org/10.1016/j.jhydrol.2020.124658.

52. Falcone, J. A. *GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow*; US Geological Survey; **2011**. https://doi.org/10.3133/70046617.

53. De Cicco LA, Lorenz D, Hirsch RM, Watkins W. *dataRetrieval: R packages for discovering and retrieving water data available from U.S. federal hydrologic web services;* **2018.** doi: 10.5066/P9X4L3GE, https://code.usgs.gov/water/dataRetrieval.

54. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; **2017**. https://www.R-project.org/

55. Cutler, Adele. Random Forests – A Statistical Tool for the Sciences; **n.d**. https://www.r-project.org/conferences/useR-2009/slides/Cutler.pdf

56. Liaw, A. and Wiener, M. Classification and Regression by randomForest. R News 2(3), 18—22; **2002**. http://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

57. Poff, N. L.; Olden, J. D.; Pepin, D. M.; Bledsoe, B. P. Placing Global Stream Flow Variability in Geographic and Geomorphic Contexts. *River Res. Appl.* **2006**, *22* (2), 149–166. https://doi.org/10.1002/rra.902.

58. Griffith, G. E.; Omernik, J. M.; Johnson, C. B.; Turner, D. S. *Open-File Report*; **2014**. https://doi.org/10.3133/ofr20141141.

59. Hill, R. A.; Weber, M. H.; Leibowitz, S. G.; Olsen, A. R.; Thornbrugh, D. J. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *J. Am. Water Resour. Assoc.* **2016**, *52* (1), 120–128. https://doi.org/10.1111/1752-1688.12372.

60. Tarboton David G.; Ames Daniel P. Advances in the Mapping of Flow Networks from Digital Elevation Data. *Bridging the Gap* **2001**, 1–10. https://doi.org/10.1061/40569(2001)166.

61. Wahl, K. L.; Wahl, T. L. BFI—A Computer Program for Determining an Index to Base Flow: US Bureau of Reclamation Water Resources Research Laboratory, **1988.** Accessed August 25, 2004.

62. Wolock, D. M. *Base-Flow Index Grid for the Conterminous United States*; pubs.er.usgs.gov, **2003,** No. 2003-263.

63. Abatzoglou, J. T.; Barbero, R.; Wolf, J. W.; Holden, Z. A. Tracking Interannual Streamflow Variability with Drought Indices in the U.S. Pacific Northwest. *J. Hydrometeorol.* **2014**, *15* (5), 1900–1912. https://doi.org/10.1175/JHM-D-13-0167.1.

64. PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 4 Feb 2004.

65. Jones, A. S., Alger, S.M.; Salehabadi, H.; Repko. A. Elasticity in the Colorado River Basin Using the Budyko Method, HydroShare; **2019.** http://www.hydroshare.org/resource/692cd36ffac24978b13b7352f62532ff.

66. Weeks In Drought https://droughtmonitor.unl.edu/Data/DataDownload/WeeksInDrought.aspx (accessed Apr 14, 2020).

67. Hall, D. K.; Riggs, G. A.; Salomonson, V. V. MODIS/Terra Snow Cover 5-Min L2 Swath 500m. Version 5. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center; **2006.** http://dx.doi.org/10.5067/ACYTYZB9BEOS.

68 .ESRI. ArcGIS Desktop. Redlands, CA: Environmental Systems Research Institute; **2016.**

69. SRTM: U.S. Geological Survey. 1 Arc-second Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection: U.S. Geological Survey; **2017.**

70. Naghibi, S. A.; Pourghasemi, H. R.; Dixon, B. GIS-Based Groundwater Potential Mapping Using Boosted Regression Tree, Classification and Regression Tree, and Random Forest Machine Learning Models in Iran. *Environ. Monit. Assess.* **2016**, *188* (1), 44. https://doi.org/10.1007/s10661-015-5049-6.

71. Larned, S. T.; Schmidt, J.; Datry, T.; Konrad, C. P.; Dumas, J. K.; Diettrich, J. C. Longitudinal River Ecohydrology: Flow Variation down the Lengths of Alluvial Rivers. *Ecohydrol.* **2011**, *4* (4), 532–548. https://doi.org/10.1002/eco.126.

72. Genuer, R., Poggi, J., and Tuleau-Malot, C. VSURF: Variable Selection Using Random Forests. R package version 1.0.4; **2018.** https://CRAN.R-project.org/package=VSURF.

73. O'Brien, R.; Ishwaran, H. A Random Forests Quantile Classifier for Class Imbalanced Data. *Pattern Recognit.* **2019**, *90*, 232–249. https://doi.org/10.1016/j.patcog.2019.01.036.

74. Vlek, H. E.; Verdonschot, P. F. M.; Nijboer, R. C. Towards a Multimetric Index for the Assessment of Dutch Streams Using Benthic Macroinvertebrates. In *Integrated Assessment of Running Waters in Europe*; Hering, D., Verdonschot, P. F. M., Moog, O., Sandin, L., Eds.; Springer Netherlands: Dordrecht, 2004; pp 173–189. https://doi.org/10.1007/978-94-007-0993-5_11.

75. Beck, H. E.; van Dijk, A. I. J. M.; Miralles, D. G.; de Jeu, R. A. M.; Sampurno Bruijnzeel, L. A.; McVicar, T. R.; Schellekens, J. Global Patterns in Base Flow Index and Recession Based on Streamflow Observations from 3394 Catchments: Global Patterns in Base Flow Characteristics. *Water Resour. Res.* **2013**, *49* (12), 7843–7863. https://doi.org/10.1002/2013WR013918.

76. Smakhtin, V. U. Low Flow Hydrology: A Review. *J. Hydrol.* **2001**, *240* (3), 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1.

77. Trancoso, R.; Larsen, J. R.; McAlpine, C.; McVicar, T. R.; Phinn, S. Linking the Budyko Framework and the Dunne Diagram. *J. Hydrol.* **2016**, *535*, 581–597. https://doi.org/10.1016/j.jhydrol.2016.02.017.

78. Beven, K. Runoff Generation in Semi-Arid Areas. *Dryland Rivers: Hydrology and geomorphology of semi-arid channels* **2002**, 57–105.

APPENDIX

Table S1. Lists of predictors selected (VSURF) for hierarchical classification models, ranked by the variable importance function in Random Forests, with notations for origin: Hill et al. [59] *StreamCat* – SC, PRISM climate models [64] – PRISM, Jones et al. [65] *Budyko* (calculated from PRISM and RunOff *StreamCat* variables) – BD/SC, and curvature tools using SRTM [69] in *ArcMap* [68]– GIS.

Watershed Attribute Description	**Abbreviation**
2-group (A, B)	
1. Base flow index for the watershed (SC)	BFIWs
2. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
3. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
4. Standard deviation of longitudinal stream (parallel stream profile–DEM.GIS)	ParCurveSTD
5. Percent of forest cover across watershed in 2001 (NLCD/SC)	PctMxFst2001Ws
6. Standard deviation of latitudinal stream curve (perpendicular plane–DEM.GIS)	PerpCurveSTD
7. MODIS terra satellite zonal statistic (MODIS/GIS)	MODISsnow
8. Compositional strength for the watershed bedrock (SC)	CompStrgthWs
9. Percent of watershed classified as barren land cover in 2001 (NLCD/SC)	PctBl2001Ws
3-group (A1, A2, B)	
1. Base flow index for the watershed (SC)	BFIWs
2. Standard deviation of longitudinal stream (parallel stream profile–DEM.GIS)	ParCurveSTD
3. Longitude of a gauge (SC)	LONG
4. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
5. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
6. Average elevation for the watershed (SC)	ElevWs
7. Average annual depth of potential evapotranspiration (BD/SC)	PET
4-group (A1, A2, B1, B2)	
1. Longitude of a gauge (SC)	LONG
2. Base flow index for the watershed (SC)	BFIWs
3. Standard deviation of longitudinal stream (parallel stream profile–DEM.GIS)	ParCurveSTD
4. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
5. Average elevation for the watershed (SC)	ElevWs
6. Average annual depth of potential evapotranspiration (BD/SC)	PET
7. Compositional strength for the watershed bedrock (SC)	CompStrgthWs
8. Mean of latitudinal stream curve (perpendicular plane–DEM.GIS)	PerpCurveMean
9. MODIS zonal statistic (MODIS/GIS)	MODISsnow

Table S1. (cont.)

5-group Selected Model*(A1, A2, B1a, B1b, B2)*****	
1. Base flow index for the watershed (SC)	BFIWs
2. Longitude of a gauge (SC)	LONG
3. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
4. Average annual depth of potential evapotranspiration (BD/SC)	PET
5. Average elevation for the watershed (SC)	ElevWs
6. The percent of grassland coverage for the watershed in 2001 (NLCD/SC)	PctGrs2001Ws
7. Percent of forest cover across watershed in 2001 (NLCD/SC)	PctMxFst2001Ws
8. Latitude of a gauge (SC)	LAT
9. Average elevation for the catchment (SC)	ElevCat
7-group (A1, A2, B1a, B1bi, B1bii, B2a, B2b)	
1. Base flow index for the watershed (SC)	BFIWs
2. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
3. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
4. Average annual depth of potential evapotranspiration (BD/SC)	PET
5. Average elevation for the watershed (SC)	ElevWs
6. Longitude of a gauge (SC)	LONG
7. The percent of grassland coverage for the watershed in 2001 (NLCD/SC)	PctGrs2001Ws
8. Evaporative index of actual evapotranspiration/precipitation (BD/SC)	ET_P
9. Percent of forest cover across watershed in 2001 (NLCD/SC)	PctMxFst2001Ws
10. Latitude of a gauge (SC)	LAT
11. Habitat provision component score calculated using watershed metrics (SC)	WHABT
12. MODIS zonal statistic (MODIS/GIS)	MODISsnow
13. Area of watershed in square kilometers (SC)	WsAreaSqKm

Table S2. Lists of ZFY and ZFD classification predictors selected by VSURF, ranked by the variable importance function in Random Forests.

Watershed Attribute Description	**Abbreviation**
ZFY 20%: 3-group	
1. Base flow index for the watershed (SC)	BFIWs
2. Maximum value of longitudinal stream curve (parallel stream profile–DEM.GIS)	ParCurveMax
3. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
4. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
5. 30-year average watershed precipitation from PRISM model (PRISM/SC)	Precip8110Ws
6. The percent of grassland coverage for the watershed in 2001 (NLCD/SC)	PctGrs2001Ws
ZFY 75%: 3-group	
1. Base flow index for the watershed (SC)	BFIWs
2. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
3. The percent of grassland coverage for the watershed in 2001 (NLCD/SC)	PctGrs2001Ws
4. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
5. Combined lat/long stream curve (profile and perpendicular plane–DEM.GIS)	ParCurveRange
6. Habitat provision component score calculated using watershed metrics (SC)	WHABT
7. MODIS zonal statistic (MODIS/GIS)	MODISsnow
8. Maximum value of latitudinal stream curve (perpendicular plane–DEM.GIS)	PerpCurveMax
9. The percent of crop land use for the watershed in 2001 (NLCD/SC)	PctCrop2001Ws
20%ZFD: 3-group	
1. Base flow index for the watershed (SC)	BFIWs
2. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
3. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
4. Maximum value of longitudinal stream curve (parallel stream profile–DEM.GIS)	ParCurveMax
5. Average elevation for the watershed (SC)	ElevWs
6. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
7. The percent of grassland coverage for the watershed in 2001 (SC)	PctGrs2001Ws
8. 30-year average watershed precipitation from PRISM model (SC)	Precip8110Ws

Table S2. (cont.)

9. Maximum value of latitudinal stream curve (perpendicular plane–DEM.GIS)	PerpCurveMax
10. MODIS zonal statistic	MODISsnow
2 & 20%ZFD: 4-group	
1. Base flow index for the watershed (SC)	BFIWs
2. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
3. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
4. Average elevation for the catchment (SC)	ElevWs
5. Maximum latitudinal stream curve (perpendicular plane–DEM.GIS)	ParCurveMax
6. Percent of watershed classified as barren land cover in 2001 (SC)	PctBl2001Ws
7. The percent of grassland coverage for the watershed in 2001 (SC)	PctGrs2001Ws
8. Percent of watershed classified as ag land cover (NLCD 2006 classes 81-82) occurring on slopes greater than or equal to 10% (SC)	PctAg2006 Slp10Ws

Table S3. List of predictors selected (VSURF) for the ZFY and ZFD Random Forests Regression models, ranked according to Variable Importance.

Watershed Attribute Description **ZFY Regression: 58% Variance Explained**	**Abbreviation**
1. Base flow index for the watershed (SC)	BFIWs
2. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
3. The standard deviation of curvature variable (GIS)	STD_curve
4. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
5. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
6. Standard deviation longitudinal stream curve (parallel stream – DEM.GIS)	ParCurveSTD
7. Compositional strength for the watershed bedrock (SC)	CompStrgthWs
8. Average elevation for the watershed (SC)	ElevWs
9. The minimum of curvature variable (GIS)	MIN_curve
10. Percent of forest cover across watershed in 2001 (NLCD/SC)	PctMxFst2001Ws
11. Average elevation for the catchment (SC)	ElevCat
Watershed Attribute Description **ZFD Regression: 57% Variance Explained**	**Abbreviation**
1. Base flow index for the watershed (SC)	BFIWs
2. Aridity index of potential evapotranspiration/precipitation (BD/SC)	PET_P
3. Average soil topographic wetness index for the catchment area (SC)	WetIndexCat
4. 30-year average watershed maximum from PRISM temp model (PRISM/SC)	Tmax8110Ws
5. The standard deviation of curvature variable (GIS)	STD_curve
6. 30-year average watershed minimum from PRISM temp model (PRISM/SC)	Tmin8110Ws
7. The minimum curvature variable (GIS)	MIN_curve
8. Percent of forest cover across watershed in 2001 (NLCD/SC)	PctMxFst2001Ws
9. Percent introduced or managed vegetation, non-native but *not* agriculture in the watershed (NLCD/SC)	PctNonAgIntrodManagVegWs